DRAFT, 1-31-1986, version b

RELAXATION METHOD OF IMPULSE DECONVOLUTION IN AN OPTICAL CORRELATOR

by

Richard D. Juday

Tracking and Communications Division
NASA Johnson Space Center
Houston, TX 77089

February 1985

(NASA-TM-89615) RELAXATION METHOD OF IMPULSE DECONVOLUTION IN AN OPTICAL CORRELATOR (NASA) 18 p Avail: NTIS

N87-70466

Unclas
00/74 0079413

DRAFT, 1-31-1986, version b

RELAXATION METHOD OF IMPULSE DECONVOLUTION IN AN OPTICAL CORRELATOR

Richard D. Juday

ABSTRACT

A method is proposed that is intended to compensate simultaneously for poorly-known diffraction patterns and imperfect elements in an optical correlator system that uses a programmable filter. One seeks a static (or otherwise easily calculated) correction for disturbances that are peculiar to a given optical correlator and are, hence, not modelable *a priori*. The technique relaxes an initial guess at a matching filter for an impulse in the input plane, in order to find filters that yield progressively more localized patterns in the correlation plane. Linearity and shift invariance of the system (or discovered departures from them) then allow construction of the matched filter for an arbitrary reference pattern. The technique is expected to be robust in the sense of having a very large capture radius and being insensitive to the form of an initial approximation to the optimal filter. The method applies to both continuous and discrete, and also to both real and complex, filters.

I. INTRODUCTION

In coherent optical correlators using spatial light modulators, the input image causes a modulation of the coherent reading wavefront (see Figure 1). The diffraction disturbance to the reading wavefront depends locally on the physics and the spatial configuration of the device, as well as depending globally on the pattern written onto the device. A filter is placed at the focal plane of the optical Fourier transform lens. In classical matched filtering, the filter has an effect equal to the complex conjugate of the Fourier transform of a reference object whose correlation with the input scene addressing the modulator is desired. The complex conjugate comprises two parameters, phase and amplitude. However, a one-parameter combination of phase and amplitude is all that can be introduced with such devices as the Litton magneto-optic device (MOD), Hughes' liquid crystal light valve (LCLV), the Texas Instruments deformable mirror device (DMD), etc. The MOD and the LCLV principally modulate amplitude; the DMD, principally phase. The challenge is to optimize a filter within the constraint of controlling only the one-parameter combination of phase and amplitude.

Modeling of phase-only filters (POF) has not yet addressed the diffraction pattern of the reading-plane pixel, nor the physically limited means of producing a desired phase modulation at the filter plane, nor the expected high sensitivity of the POF correlator to phase deviations of the order introduced by real optics. Horner and Gianino [1] indicate a scheme which will use an iterative technique to optimize a POF derived from a classical matched filter, but the details have not been published. In this paper there is presented an iterative, adaptive filter-optimizing method that is independent of the starting filter and conforms to the control constraints.

II. SOME PRACTICAL DIFFICULTIES IN OPTICAL CORRELATORS

The pattern brought to the filter plane of even a linear space-invariant optical correlator system is not exactly the Fourier transform of the original input image, but instead is the image's convolution with the system's transform of the diffraction pattern of a single pixel in the reading plane. Imperfections existing in any physical optical system interfere with the purity of the optical Fourier transform. These include non-planarity of surfaces, scattering centers such as dust or inclusions, and imperfect optical alignment. It takes little imperfection in a high-index transmitting element to cause appreciable departure in phase from what perfect optics would have given. Altogether, the filter plane may be presented with a complex wavefront differing significantly from the Fourier transform of the original input image. If the input image is the desired reference object, the one-parameter spatial light modulator (SLM) is asked to match both phase and amplitude of the arriving wavefront, so as to produce a bright spot in the correlation plane (after passage through subsequent imperfect optics). This will likely be difficult, particularly for phase-only filtering, since the difference from perfect Fourier transform behavior induced by, for example, a two-wave-flat beamsplitting cube will be far more significant in phase than in amplitude. Furthermore, the physical configuration of the filter may not fit the shape of the transformed input (particularly under the constraints of one-parameter control) even if that were known exactly. Modeling of phase-only matched filtering [2,3] shows that the POF has the possibility of producing very sharp correlation peaks; concomitantly one expects that the POF will be very sensitive to phase errors in the filter (which would arise as uncertainty in the diffraction pattern of an individual reading-plane pixel or as non-uniformity of phase propagation by the physical optical system).

An amplitude-only filter (AOF) produces broader correlations than the POF [3]; the hypothetical two-wave-flat beamsplitter affects the transformed wavefront's amplitude less strongly than its phase; and measurements of amplitude (intensity) at the filtration plane are easier than measurements of phase. Consequently, producing an optimized filter is expected to be more difficult for the POF than for the AOF. However, the promise of the POF's strong signal-to-noise ratio justifies an attempt to realize its potential, and the easier AOF can also benefit from the method.

III. THE METHOD OF RELAXATION BY GRADIENT SEARCH

A method of finding the optimal filter in a physical system has been devised. It will allow investigating the departures of the correlator system from the desirable qualities of linearity (superposition) and shift invariance. The method is similar to adaptive telescope optics in performing adjustments on the filter to produce the sharpest spot in the correlation plane, even as adaptive telescope optics sharpens a star's image. Adaptive telescope optics adjusts comparatively few parameters at a high rate, to accommodate temporal instabilities; the adaptive relaxed filter adjusts many parameters once and assumes temporal stability.

The technique is a gradient search among all possible variations from the initial guess at a proper filter. We need two tools: the scalar whose value is to be maximized, and a complete and appropriate space in which to take the gradient of the scalar. The more accurate the starting filter, the more rapid is convergence to the optimal filter, though the method is designed to have a final result that is independent of the exact form of the starting filter.

For an exhaustive search in a space of high dimension, one requires a rapidly calculable metric for the sharpness, S. The metric must avoid capture by false (local) maxima. To meet these goals, a sequence of sharpness measures is envisioned with the measure at any one time tailored to the characteristics of the current pattern. The sequence begins with a low spatial resolution and proceeds to higher and higher resolution. The ideal final measure is the intensity of light at the one pixel in the correlation plane which corresponds to the location of an input plane impulse. If the light is initially distributed over all the correlation plane, starting with such a sharp metric almost guarantees being caught by a local maximum. The technique proposed here is to sum the light within regions of progressively diminishing size; by analogy, if one has a given volume of sand scattered on a plane, the tallest pile of sand is most economically constructed by beginning inward radial sweeping at the edges (rather than at intermediate radii) and by steadily reducing the radius at which one sweeps. The succeeding radius in a sequence could be, for example, a fixed fraction of the radius of gyration of the present intensity pattern, and we would take the metric of sharpness, S, to be the amount of light inside the current value of the sensing radius.

We can apply the method to both discrete and continuous-valued filters with only slightly different procedures. Let us first examine the continuous filter.

Having determined a metric of sharpness as the scalar to be maximized, we need the space in which we will calculate sharpness as a function of position. The space of Hadamard coefficients is appropriate. Brief mention of certain qualities of the Hadamard basis is appropriate. We follow Pratt [4] in this development.

The Hadamard transform is based on the Hadamard matrix, a square array of plus and minus ones whose rows and columns are orthogonal. There are N^2 orthogonal matrices that are N on a side, so they are complete with respect to representing any N x N pattern. The matrices are used as the basis with which to represent a square array of numbers -- in the present case, the control values applied to the spatial light modulator. (If we were controlling both phase and amplitude at each grid location in the SLM we would have complex coefficients for the Hadamard basis planes, and without loss of generality we can continue using real scalars for the coefficients.) The Hadamard basis can be normalized by a single multiplicative scalar for each matrix. The two-dimensional matrices can each be expressed as the product of two one-dimensional functions, with important ramifications for economy of generation of the basis planes. The set of N^2 2-D basis planes is generated by the outer product of pairs drawn from the N 1-D functions. The 1-D functions can be arranged in order of sequency (the number of sign changes in the function; sequency is closely related to spatial frequency), and so the 2-D basis planes can be ordered in two dimensions of sequency. The N x N Hadamard basis is easily calculated where N is an integer power of 2, a very handy property in light of the fact that most spatial light modulators come in arrays that are powers of 2 on a side. The orthogonality of the Hadamard matrices leads easily to the representation of a given 2-D array by the Hadamard coefficients. If we desire to represent $t(j,k)$ by the set of coefficients $a_{u,v}$ (which is the same as the transform of the array) the orthonormality of $H_{u,v}(j,k)$ easily yields

$$t(j,k) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} a_{u,v} H_{u,v}(j,k) \qquad (1)$$

where

$$a_{u,v} = \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} t(j,k)\ H_{u,v}(j,k) \tag{2}$$

Pratt[4] gives another method that operates on the bit patterns of the transform coordinates; letting $t(j,k)$ be the one-parameter control applied as the filter, its Hadamard transform, $T(u,v)$, is efficiently calculated by:

$$T(u,v) = \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} t(j,k)\ (-1)^{q(j,k,u,v)} \tag{3}$$

where

$$q(j,k,u,v) = \sum_{i=0}^{N-1} [g_i(u) j_i + g_i(v) k_i] \tag{4}$$

and

$$\begin{aligned} g_0(u) &= u_{n-1}, \\ g_1(u) &= u_{n-1} + u_{n-2}, \\ g_2(u) &= u_{n-2} + u_{n-3}, \\ &\ldots, \\ g_{n-1}(u) &= u_1 + u_0 \end{aligned} \tag{5}$$

and u_i is the state of the i^{th} bit in the binary representation of u. For example, if $u=13$, then $u_3=1$, $u_2=1$, $u_1=0$, and $u_0=1$. Logical bit operations in computation machinery make this an efficient and fast method of representing the filter.

The transform $T(u,v)$ of a filter can be regarded as the components of a vector of length N^2 by column- or row-scanning $T(u,v)$. Define $w := Nu+v$; the relationship is invertible since u and v both run from 0 to (N-1). The transform $T(u,v)$ scans into a vector $\underline{F}(w)$ by $\underline{F}(w) := T(u,v)$. The Hadamard basis being complete for arrays N x N, then $\underline{F}$ is able to represent any filter in dimension N^2. Since $\underline{F}$ gives T(u,v) which in turn gives the filter, the information in $\underline{F}$ is the same as the information in the filter itself. We search on values of $\underline{F}$ for the one giving the sharpest output in the correlation plane of the optical correlator.

We begin by placing an impulse at a single location in the input plane and making an estimate of its matched filter (AOF, POF, or other as appropriate). Next we represent the filter with respect to the Hadamard transform basis functions. Then the sharpness S is obtained for that vector, and the components of $\underline{F}$ are adjusted to maximize S. Beyond the usual properties of having row and column orthogonality and being complete, the Hadamard basis offers the advantage that the magnitude of each element is the same. As any component of $\underline{F}$ is adjusted, every pixel in the filter is affected with the same magnitude; the visibility of the effect will be larger than if one or only a few filter pixels were changed. The possible drawback of the high sensitivity to change of each coefficient is that the simple gradient search outlined here might have to be supplemented by an interpolation scheme. We need a hill-climbing algorithm and will outline one such to find the largest value of S .

Let $\underline{F} = (F_i)$ be the vector of Hadamard coefficients. Let $S_j[\underline{F}]$ be the j^{th} measure of sharpness; let S_1 be the total amount of light recorded at the correlation plane, and let S_j converge toward S^*, the amount of light recorded at only a single central pixel.

THE ALGORITHM

For j in its domain do the following:

For i in its domain do the following:

Using suitable small values of ϵ and β find the maximum of $S_j[\underline{F}]$ on the i^{th} axis as follows.

Estimate the i^{th} component $\nabla_i S_j[\underline{F}]$ of grad(S_j) by

$$\nabla_i S_j[\underline{F}] = \frac{S_j[\underline{F}+\epsilon \hat{h}_i] - S_j[\underline{F}]}{\epsilon} \qquad (6)$$

in which $\hat{h}_i$ is the unit vector along the i^{th} axis in the N^2-dimension space of Hadamard coefficients. Replace $\underline{F}$ with $(\underline{F} + \beta \hat{h}_i \nabla_i S_j[\underline{F}])$. Repeat until the maximum is found; there $\nabla_i S_j[\underline{F}] = 0$.

Repeat through all values of i.

If the sharpness $S_j[\underline{F}]$ increases as a result of an i-loop manipulation, repeat the i-loop with that value of i.

Here we have maximized sharpness for the j^{th} measure S_j; further manipulations on i do not increase S_j. Thus we repeat with the next measure of sharpness until we have used all values of j. Let the final value of $\underline{F}$ be designated $\underline{F}^+$. Repeated cycling through the diminishing radii will have the effect of peristaltically moving light energy into the central correlation spot.

We have not shown that $\underline{F}$ necessarily converges to $\underline{F}^*$, which of all possible values of $\underline{F}$ maximizes S^*. Indeed, given the sharpness and noisiness shown in the modeling of Horner and Gianino[3], one can imagine that even <u>starting</u> with $\underline{F}^*$, proceeding to use the sequence of S_j's as above would lead to a value $\underline{F}^{+}$ such that $S^{*}[\underline{F}^{+}] < S^{*}[\underline{F}^{*}]$. However, we are certainly no worse off for the effort; if we find that $S^{*}[\underline{F}^{+}] < S^{*}[\underline{F}_1]$, with $\underline{F}_1$ the starting filter, we can always use $\underline{F}_1$.

The sizes of ϵ and β will depend on various considerations related to convergence, sensitivity, and physical limitations. The correlation peak may not be detected, depending on ϵ and β, as the light is swept toward the desired center. The size of ϵ is chosen small enough to avoid stepping beyond a nearby maximum but large enough to give a significant change in S , which allows estimation of the gradient component. ϵ and β are also chosen within limitations imposed by the physical nature of the SLM; there are maximum and minimum activations for each pixel. Adaptations of the procedure will have to be made for individual instances. As an example, if there is a substantial DC term in the transform plane due to a significant proportion of inactive area on the input SLM, then the sharpness measure could be taken as the difference between the correlation plane patterns with the impulse at the input plane and its current estimated filter both on and both off. The differencing decreases the effect of the unchanging large-amplitude portion of the signal.

Now we turn to the discrete filter. The binary phase-only filter (BPOF) will serve as an example. If it were possible to represent an arbitrary N x N pattern of plus and minus ones as, say, the element-wise product of a subset of the Hadamard matrices, then the analog of the gradient search outlined above could be a search for the minimal subset of the Hadamard matrices to multiply element-wise for the optimum filter. Advantage of the binarization would be taken in there possibly being fewer than N^2 parameters required to represent a filter. Unfortunately it is not generally possible to do such a product representation. The BPOF can, however, be regarded as simply a thresholded continuous filter. The value of the threshold is one more parameter in the representation of a filter by the Hadamard coefficients, and it thus it is one more coordinate for the gradient search. The principles are otherwise as stated above, though it is recognized that since there is a many-to-one aspect to this representation of a binary filter, there will be areas of apparently zero gradient in the search.

However, there is presumably a "best" BPOF. The Hadamard basis is exhaustive in representing patterns (including binary ones and hence the best BPOF). Given a continuous filter and a binarization threshold that together give exactly the best BPOF, there will be a more or less graceful degradation from the best BPOF as the N^2+1 parameters are altered. One hopes that the inverse operation would work -- that convergence to the best BPOF would ensue.

This is more an "existence" argument than a recipe; the limitation to binary values should admit of a computationally economic representation for gradient search, and research into such economic representation will continue.

The extension to ternary filters is done by the extension to two thresholds versus one, and so forth for any number of levels in a discrete filter.

IV. LINEARITY, SHIFT-INVARIANCE, AND ARBITRARY REFERENCE IMAGES

Once the optimal filter F^* has been found for an impulse at one location in the input plane, the putative linearity and shift invariance of the optical Fourier transform are examined. We hope for linearity in the sense of proportional response in the strength of the impulse, and shift invariance in the sense of the usual phase effect from a shift of origin in the Fourier transform. A truly linear system will also show superposition; the optimum filter pattern for pairs of impulses would be the sum of their individual filters. Assuming such qualities leads to starting filters for different locations of the input impulse, etc. One would expect that the search for the optimal filter of the shifted impulse would be less strenuous than finding the original impulse's filter; the technique is the same but the starting point should be nearer the final value. Consistent departures from the shift invariance, as due to to imperfect optics, are sought and modeled. Similar comments apply for examinations of linearity (proportionality) and superposition.

If linearity, shift invariance, and superposition hold (or if departure from them is tractable), then an arbitrary reference image can have its matched filter created from the filters for the impulses that sum to the arbitrary input image. One hopes to find static -- or at least easily computed -- corrections to the transform of an arbitrary reference image. Otherwise, the system would be trained by the relaxation technique to recognize each new reference image. There is an obvious preference for the former.

The technique is a quite general one, applicable to any sort of rapidly programmable filter. Rapid programmability is a requirement if the method is not to get out of hand, but if the procedure is automated and done efficiently it will not necessarily be too time-consuming. For example, at the 30Hz which is not an unreasonable speed for reading new patterns into a filter, it takes just over nine minutes to count through the 16K Hadamard patterns required to represent a 128 x 128 filter. The technique will be tested on an optical correlator system being developed for NASA's Johnson Space Center by Texas Instruments. The Texas Instruments DMD will be used for both the reading plane and the filter plane in the correlator. Results will be reported later.

REFERENCES

1. J. L. Horner and P. D. Gianino, Appl. Opt. 24, 851 (1985).
2. Joseph L. Horner and James R. Leger, Appl. Opt. 24, 609 (1985).
3. J. L. Horner and P. D. Gianino, Appl. Opt. 23, 812 (1984).
4. William K. Pratt, Digital Image Processing, Wiley, New York, (1978).

DRAFT, 1-31-1986, version b

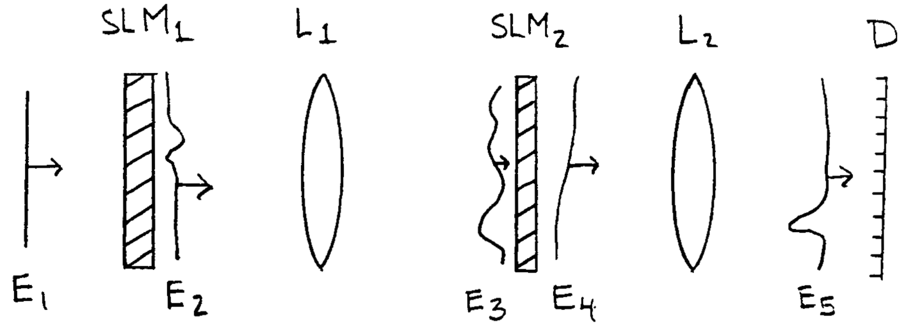

Figure 1. An optical correlator laid out in simplified all-transmitting form. The coherent wave E_1 becomes E_2 after reading the information impressed on the spatial light modulator SLM_1. Lens L_1 transforms E_2 to E_3 at the filtering SLM_2. Passage through SLM_2 results in E_4 which is transformed by L_2 to E_5 at the imaging detector D. A pattern put in at SLM_1 is "matched" by the pattern on SLM_2 that produces the most localized and centered pattern at D.